

Beating Cheating: Dealing with Collusion in the Non-Iterated Prisoners Dilemma

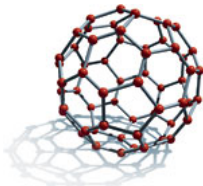
N. Höning, T. Kozelek, M.C. Schut

October 30th, 2008



Problem Domain

In complex and dynamic markets, most of the opponents have never interacted before and it is hard to decide for them whether they can trust each other.



We propose that **agents can ask other agents they already trust if the current opponent can be trusted.**



Collusion

Trust schemes might be rigged by collusion.



We want to tackle that, too.



- 1 Background
 - Problem Domain
 - Concepts
- 2 The Ask-First Strategy
- 3 World Model
- 4 Results
 - Setup
 - Graphs



The Prisoner's Dilemma

- A non-zero-sum game with two players
- **C**ooperate (long-term) or **D**efect (short-term benefit) ?

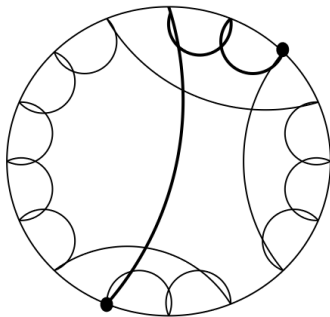
	A	B	payoff: A	payoff: B	Σ
traitor benefit (t)	D	C	5	0	5
reward (r)	C	C	3	3	6
punishment (p)	D	D	1	1	2
sucker payoff (s)	C	D	0	5	5

- iterated/non-iterated



Small World Networks

Our world is strongly interconnected ...



- Short average path lengths ("six degrees of separation")
- High clustering index (# of direct neighbours higher than random graphs)



2. The Ask-First Strategy



The *Ask-First* Strategy

If cooperation would be good in the long run (PD), and if everyone is well-connected to everyone (SW) - then gossip might be informative for cooperators.



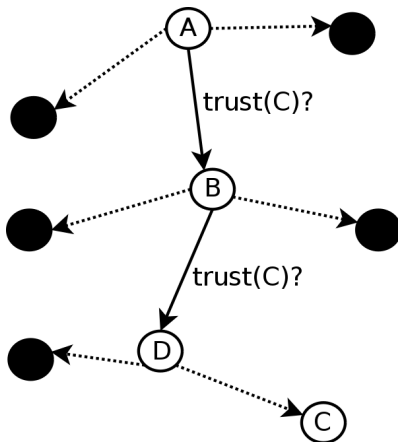
The *Ask-First* Strategy

Ask-First Agents *ask* for advice on opponent before they *act*.

- select best neighbour
- ask for advice about opponent
- use the advice: Decide whether to cooperate or not



Managing Trust



B chooses D via:

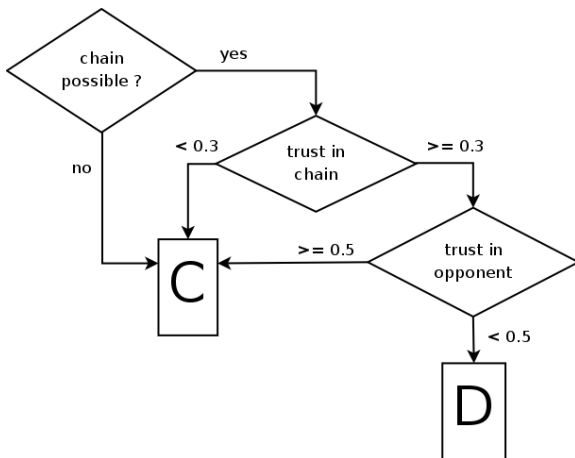
max :

$dist_bias * dist(D, C) +$
 $trust_bias * trust(B, D)$

A's trust in the chain:
 $trust(B, D) * trust(A, B)$



Decision Making



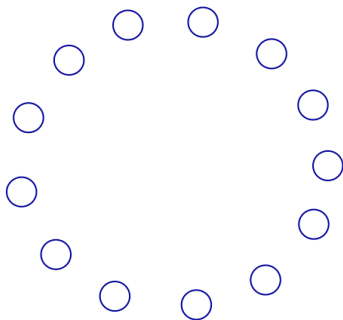
Ask-First Agents are cooperators per default.



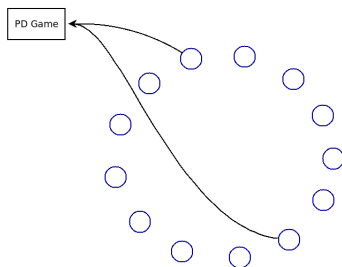
3. World Model



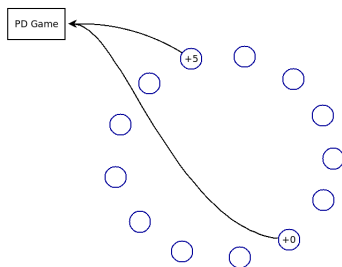
a decentral structure



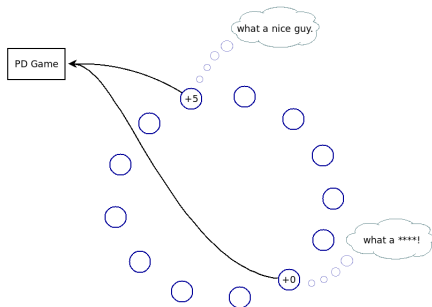
random agents play the Prisoner's Dilemma



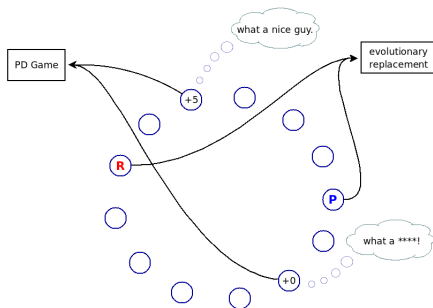
after each game, agents receive payoff



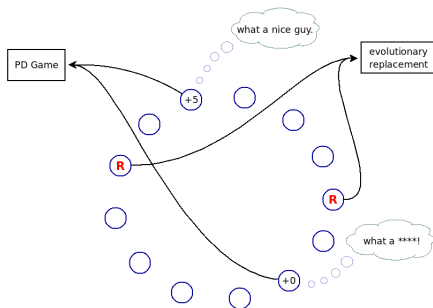
agents remember a trust value for previous opponents



evolutionary updates are done with two random agents



the poor agent is replaced with a copy of the richer one



4. Results



Hypotheses

- 1 The AF strategy is successful against defection.
- 2 The AF strategy is successful against collusion.
- 3 Colluding agents cannot spread false information over the information chains.



Strategies

- *Always Cooperate (AC)* - cooperates in every interaction.
- *Always Defect (AD)* - always defects the opponent.
- *Ask First (AF)* - asks before it acts
- *Simple Collusion (SC)* - always defects, lies about trust in others ($t = 1 - t$)



Measurements

X-Axis:

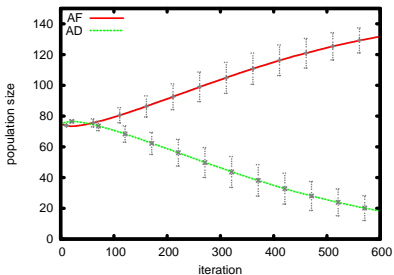
- iterations

Y-Axis:

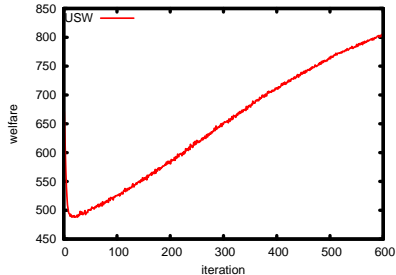
- Population Size
- Utilitarian Social Welfare
- Chain usage



Hypothesis 1: The AF strategy is successful against defection.



(a) Strategy populations

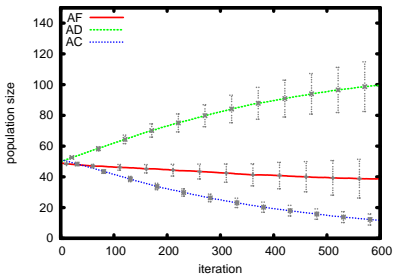


(b) Utilitarian Social Welfare

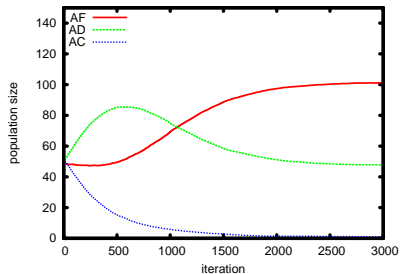
AD vs AF in experiment 1, N=150



Hypothesis 1: The AF strategy is successful against defection.



(c) Strategy populations

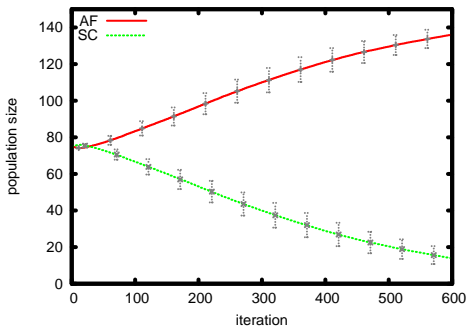


(d) Strategy populations (3,000 iterations)

AD vs AC vs AF in experiment 2, $N=150$



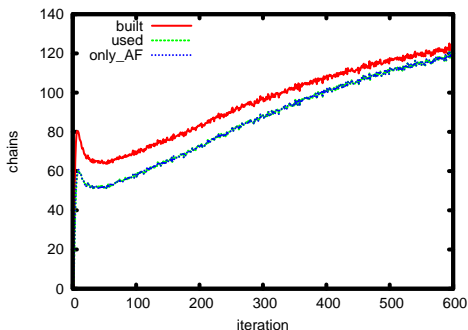
Hypothesis 2: The AF strategy is successful against collusion.



AF vs SC in experiment 3, N=150



Hypothesis 3: Colluding agents cannot spread false information over the information chains.



AF vs SC in experiment 3, N=150



Summary

- a gossip-based, decentralised trust system
- successful for cooperators
- the mechanism excludes colluders

